An Expert System for Weapon Identification and Categorization Using Machine Learning Technique to Retrieve Appropriate Response

Rana Mohtasham Aftab[1], Mariam Ijaz[2], Dr. Faisal Rehman[3], Muhammad ishtiaq[4]

[1,2,3,4] Department of Computer Science, Lahore Leads University, Pakistan

Email: ranasabh@hotmail.com

ABSTRACT

In response to any terrorist attack on hospitals, airports, shopping malls, schools, universities, colleges, railway stations, passport offices, bus stands, dry ports and the other important private and public places, a proper plan will need to be developed effective response. In normal moments, security guards are deployed to prevent criminals from doing anything wrong. For example, someone is moving around with a weapon, and security guards are watching its movement through closed circuit television (CCTV). Meanwhile, they are trying to identify his weapon in order to plan an appropriate response to the weapon he has. The process of manually identifying weapons is man-made and slow, while the security situation is critical and needs to be accelerated. Therefore, an automated system is needed to detect and classify the weapon so that appropriate response can be planned quickly to ensure minimal damage. Subject to previous concerns, this study is based on the Convoluted Neural Network (CNN) model using datasets that are assembled on the YOLO and you only see once. Focusing on real-time weapons identification, we created a data collection of images of multiple local weapons from surveillance camera systems and YouTube videos. The solution uses parameters that describe the rules for data generation and problem interpretation. Then, using deep convolutional neural network models, an accuracy of 97.01% is achieved.

KEYWORDS: CCTV, CNN, YOLO, WDR, Real Time, Detection, Categorization

1. INTRODUCTION

According to the World Health Organization, the rate of loss of human life due to barbaric crimes in a year is 15,000 and these crimes can be committed using cold steel weapons, for example axes, Swords and knives etc. One of the reasons for such barbaric illegal activities in different parts of the world is that the legal possession of a gun for a short period of time, even later, in such areas, has been outlawed [1]. According to shared facts and figures from the United Nations Office on Drugs and Crime (UNODC), gun crime is on the rise in every 100,000 inhabitants in many countries. Psychologically, it has been proven that if a person has a weapon of any kind, he can use it to commit a crime. Due to the increase in the world's population, criminal activities are increasing rapidly. Improving traditional security practices takes time. The reaction of the traditional policing system leads to investigations after incidents such as robbery, snatching and violence. However, these methods are insufficient to prevent these violent incidents [2]. In the modern world, technology is the basic structure of public and national security. Close circuit television (CCTV) camera surveillance and control systems have been installed around the world to monitor such incidents, but it requires human resources to identify the situation. Therefore, continuous monitoring in the surveillance camera system is flawed and cannot be done every second. Therefore, there is a need for an automated surveillance system that can detect a weapon directly from the scene and alert security personnel to take action to prevent any violent activity [3].

Previously, most studies have identified the detection of weapons using classical machine learning methods using RGB images. Currently, deep learning models are the most accurate identification models based on the Region Proposal Network (RPN). Base models are the most accurate, improving accuracy while reducing computational costs. Faster RCNN and RFCN are the most popular CNN models, surpassing every other classic machine learning detection method with accuracy and speed [4]. Many researchers are working on real-time weapons detection systems aimed at reducing illegal activities that are increasingly using RCNN to address this issue in video. Indicate The research work focused on pistols only and was used to diagnose movie clips. Previous research has mostly been done on datasets containing fiery images downloaded from the Internet and movie images. Most of the research was done to detect weapons from CNN models (faster RCNN and RFCN). There are several challenges in weapons detection:

- Different shadows and lightning effects in pictures.
- Involvement in images that conceal weapons information.

- Position and size of weapons.
- Techniques of weapons handling.

The focuses of this study is on accurate and rapid detection of the most common types of weapons used in crime such as machine guns assault rifles, pistols, shotguns, revolvers and assault rifles. The main contribution of this research work is to present a model with good accuracy on a newly developed dataset. This work contributes to some aspects which can be summarized as follows [5].

- The parameters for generating the problem dataset are specified.
- Developed a new labeled dataset
- Rules outlined for our dataset labeling
- Develop and train models with maximum polling for detection and classification.

2. LITERATURE REVIEW

CCTV cameras are playing an important role in security and surveillance purposes. Gustav Alexandri et al. In 2017 studied the random incidents captured by CCTV cameras and inspected CCTV surveillance cameras on illegal activities. The authors found an increase in the crime rate (24% - 28%) and this could be reduced with the installation of CCTV cameras. Their study shows that CCTV cameras can reduce crime rates in many ways [6].

Alberto Castello et al. in 2019 Introduces brightness guided pre-processing for automatic detection of cold steel weapons in CCTV camera videos using deep learning. The purpose of their work was to automatically detect cold steel weapons in lighting conditions through an automatic alarm system. For their work, the authors have analyzed the region's selection techniques and the long-awaited collections of CNN [7].

Roberto Olmos et al. in 2019 introduced an approach to false-positive mitigation and called it binocular image fusion. In his study, he presented an image fusion method to develop an object detection model to find the purpose of the action that takes place in this scene. He developed a dual-camera system that was a low-cost balance for feature map computing and exploited this knowledge to improve the selection of areas for input frames. He concluded from the results of the system that his proposed style has reduced the false positive advertising which has also improved the overall production of the model for object detection [8].

Xiongwei Wu et al. in 2019 gave a brief overview of recent improvements in object detection using deep machine learning. He conducted a large survey of deep learning detection models and described them in three major sections: the detection component, learning strategy and application, and benchmarks. They also described the element which enhances the performance of the detected model. He also pointed to the future, including big batch learning, incremental learning for which the tracking model has forgotten itself destructively [9].

Ronghua Luo, Huailin Huang, Weizeng Wu. in 2020 Evaluated the discovery of significant objects based on the spinal cord extension network. He described previous research perspectives on backbone networks, most of which used pre-trained models and retrained on new data for better work results. Weakness usually results in a tingling sensation. He proposed a framework with a backbone network. To get a better and more diverse feature, they used an encoder with a dual backbone network structure. He reviewed his views on six public datasets [10].

Gyanendra K. Verma et al. presented a research to detect automatic weapons such as a gun or a pistol by using a convent or CNN in a video rotation. In their research, they used the Deep Conventional Network (DCN), a revolutionary model: the F-R CNN model, through transfer learning. He concluded that CNN performs interclass variations and interclass similarities. Different parameter values are used to cross-validate the system for system evaluation. This system can distinguish between different types of weapons such as the pistol, revolver instantly and through the scale or balance, affine, rotation or movement, and strong in the presence of different. It is worth mentioning that the accuracy of the system is 93.1%[11].

Justin Lai et al. described that their method addresses the issue of real-time detection. For the baseline model, they used a pre-trained classification model based on VGG-16 on the ImageNet dataset. The ImageNet Object Detection Dataset contains approximately 1.3 million images with approximately 1000 object classes. They create their own dataset of 3000 images based on weapons weapon and gathered from IMFDB (Internet Movie Firearm Database), home-made videos, and news frames based on selected surveillance or YouTube videos are collected. They downloaded 3,000 images of the object detection model training set and 500 images for verification compiled from the Internet Movie Firearm Database, and took 2535 images for the training and 218 images for the test. It gave excellent results on training with an accuracy of 93 % [12].

Dr. Virendra. V. Shete et al. explained how CCTV cameras are gaining importance day by day due to modernization. Their work represents a monitoring system that implements motion detection algorithms with dangerous features. He also described an emerging technology called the Internet of Things, or IoT. They implemented a motion detection video from the cameras that would be processed in a raspberry Pi and the algorithm would be able to determine if the motion of the moving object was of interest. They enhanced the system by incorporating the ability to recognize human beings in the given video. System record at 640x480 resolutions at 24 fps. Their technique is 80% accurate when detecting humans [13].

Garima Mathur et al. explain what monitoring means. To encourage, guide and protect public administration by closely monitoring and monitoring the public's actions, conduct and other information related to people's goods. They surround themselves with everything possible to explain surveillance. Timelines Background Earnings, based on semantic analysis, outlined approaches to a boasting solution to the presence of difficulty with 95% accuracy results. They presented the results of their review on a form or multiple findings, which contained methods and techniques for resolving discrimination issues related to research. He also described his merits and demerits and the scope of his future work [14].

Shaoqing Ren et al stated that the state-of-the-art network for object detection relies on Algos to propose the location of the object. Developments such as SPPNet and Fast RCNN have reduced the running time of these tracking networks due to exposure to the region's proposal computation. In their research, they claim to have presented a region proposal network that binds complete image convolutional topographies with the results of the detection network in the region proposal, which is almost free of cost. They said that at the time of testing his technique, almost all selective computational burdens are relinquished. Their detection system has also been developed on commercial systems that include user feedback such as Interest. He claims in the paper that the results show that his technique has a cost-effective environment for practical use and to improve the accuracy of the detection of objects [15].

Jifeng dai et al. described that the deepest network for object detection can be identified in all networks by the layer of interest. The first area is free from the sub-network of interest and the second area is the sub-network of interest in which the counting of shares is not allowed. In their research, they found a completely convolutional, organized and region-based object. They claim that their area-based detectors are fully convolutional, with the whole picture being based on a region like the combined FRCNN or Faster RCNN and costing a region hundreds of times at all costs. For their research, they used RFCN as a region-based fully conventional neural network. They used a two-step object detection method. First region proposal and second region classification. They used Resnet 101 to compare to the faster R-CNN, which is currently the hardest and best performer on PASCAL VOC [16].

3. METHODOLOGY

The proposed system is a model for the detection and categorization of firearms for national and public security that facilitates active security measures. This section details the proposed method of improving automatic firearms detection and categorization systems. The full flow of the proposed model for identification and classification of firearms is shown in Figure 4.1. The proposed procedure is divided into the following sections:

- The data collection parameters that are used to collect the data and then the training data are described on the rules that we have explained for our problem.
- All images from the dataset have been converted to the appropriate dimensions.
- The data is interpreted because we have introduced our dataset.
- Our proposed model was trained in proposed dataset.
- The results were prepared by providing test images to test our model [17].

3.1. Yolo Network Architecture

YOLO is known for its high accuracy and the ability to run continuously. The calculation in this picture "only looks once" as it requires only one further extension through the neural organization required for expectations. After pressing as much as possible, articles are recovered along with its bound boxes. Along with Yolo, a lone CNN simultaneously predicts multiple boundaries and class possibilities for these matters. Yolo puts trains on full images and paves the way for legal approval. This model has numerous advantages over other object recognition strategies such as it is surprisingly fast, it sees the whole picture during preparation and testing so the relevant data about the classes with their appearance Encodes with authentication. In addition, Yolo also learns to photograph articles in general with this article when, when drawn on regular images and craftsmanship is attempted, the calculation kills another technique of location. The complete network architecture of the YOLO V3 is shown in Figure 1.

Figure 1: YOLO V3 Network Architecture

3.2. Data Gathering and Augmentation

The function of deep learning neural networks improves with the measurement of available information. Data augmentation is a strategy to create new data from existing data. The data used in this research work includes 7,000 images and the dataset is divided into test dataset, training dataset, and validation dataset with a fraction of 30% and 70%, respectively. Dataset pre-processing is also done in Which the images of the dataset are converted to a normal scale of 1280x738 split. The data are interpreted to define the criteria and parameters of YOLO. The dataset contains various images obtained from a variety of sources. One image contains at least one gun and the maximum number of guns used in this research. The proposed method is presented in Figure 2, and the sample images show a variety of locally available weapons, such as Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7.

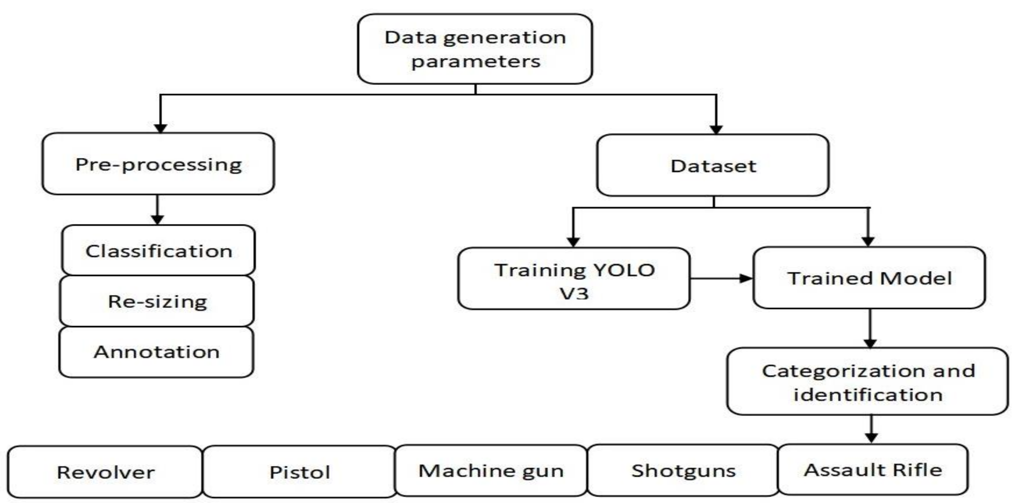

Figure 2: Methodology

4. RESULTS AND DISCUSSION

This section examines the proposed system, firearms detection and categorization system. We have extensive experience in training and testing our firearms detection and categorization systems. We have tested the proposed system following the principle of 70-30 split rule. The local dataset contains

7,000 images of local weapons. We have used DarkNet Yolo for training our models and have found encouraging results.

Figure 3. Sample Input Image 1

4.1. Confusion Matrix

A Confusion Matrix of the binary classification is a two-by-two table containing the number of four binary ranking results, usually named TP, FP, TN, and FN, instead of the "real positive number." Is interpreted as . The values of TP, FN, FP and TN are 582, 26, 11 and 622 respectively.

Figure 4: Sample Input Image 2

Figure 5. Sample Input Image 3

4.2. Performance Evaluation Matrices

Performance appraisals are based on accuracy, precision, F1 measurement and recall. To achieve encouraging results, we have focused on the challenges faced by previous models such as different sizes, shapes, distances, image angles, light effects, image resolution, imaging images, shadows and many more of the weapons. Some. We used a dataset with 7,000 images.

4.2.1. Accuracy

Accuracy is the number of data points that accurately predicts out of all data points. More formally, it is defined as the number of true positives and true negatives, divided by the number of true positives, true negatives, false positives, and false negatives.

Accuracy = TP + TN / TP + FP + FN + TN

Proposed Model = 582 + 622 / 582 + 11 + 26 + 622 = 1190 / 1352 = 0.9701

4.2.2. Recall

In pattern identification, retrieval of information, and classification, remember, is only a small part of the total amount of related events that were actually retrieved.

RECALL = TP / TP + FN

Proposed Model = 582 / 582 + 26 = 582 / 608 = 0.9572

4.2.3. Precision

In identifying patterns, retrieving information and classification, precision is the fraction of all the relevant instances which are among the retrieved instances.

Precision = TP / TP + FP

Proposed Model = 582 / 582 + 11 = 582 / 593 = 0.9814

Figure 6: Sample Input Image 4

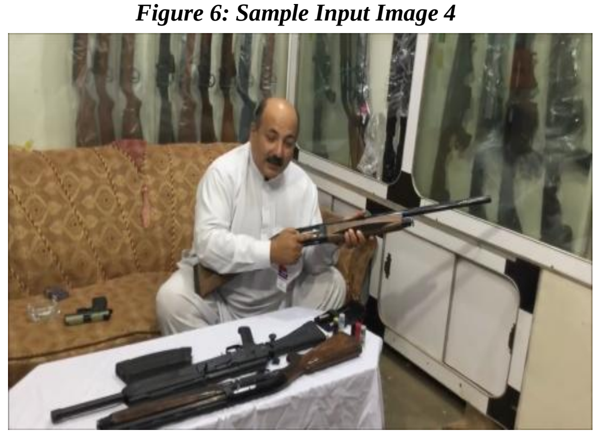

Figure 7: Sample Input Image 5

4.2.4. F1-Score

The purpose of F1 score is to transmit balance between the recall and precision.

F1-score = 2 (Recall Precision) / (Recall + Precision)

Proposed Model = 2 (0.9572 0.9814) / (0.9572 + 0.9814)= 1.8787 / 1.9386 = 0.9691

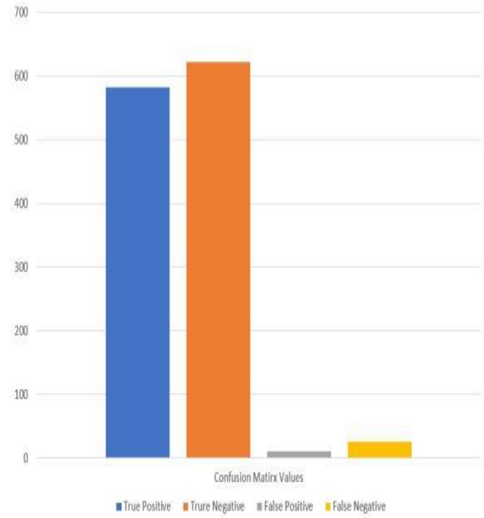

Figure 8: Confusion Matrix Graph

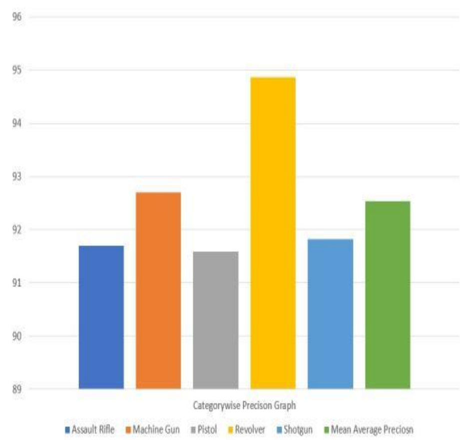

Figure 9: Precision Graph

4.3. Graphical Representation of Precision and Loss of Proposed Model

We have represented the accuracy and loss of training of the proposed model up to 5500 repetitions to see how our graph turns into its maximum value. In addition, the developed dataset and the trained model are tested on the dataset. Some of the results are shown in Figure 11, Figure 12, Figure 13, and Figure 14 and with the precision graph in Figure 9 and the deficit graph in Figure 10.

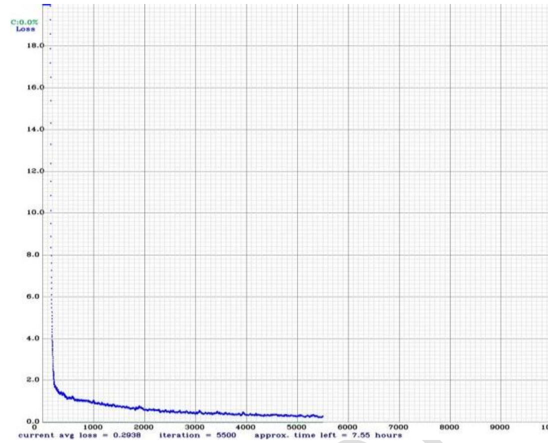

Figure 10: Loss Graph

5. RESULTS

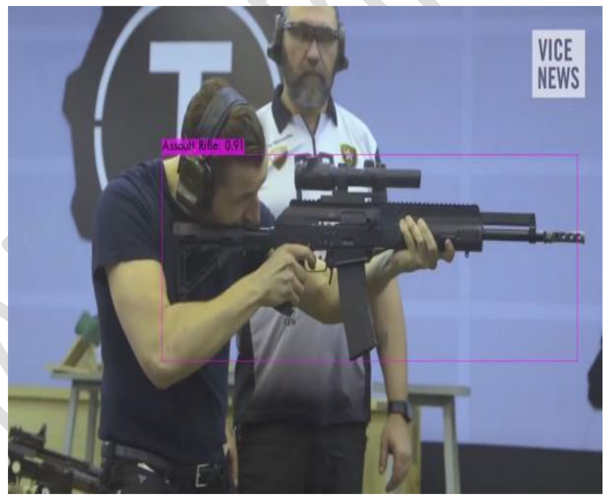

Figure 11: Results 1

Figure 12: Result 2

Figure 13: Result 3

Figure 14: Result 4

6. CONCLUSION

In this work, we presented a system that integrates deep convolutional network models. Most of the detection of objects is done in modern times by CNN-based neural networks. The main motive behind this CNN-based firearm detection and classification system is to create a virtual neural network to reduce false positives and create a model with real-time capabilities that primarily attract the attention of researchers. For baseline CNN or feature extractor, we have used Darknet YOLO. We trained this model and found the results. We have used pre-trained feature extractors as it saves a lot of time in fixing a model according to our problem. Experiments show that our model has achieved good results with 97.01% accuracy for the firearm identification. In the future, we will increase the number of categories and collect large datasets using state of art.

REFERENCES

[1] G. Alexandrie, "Surveillance cameras and crime: A review of randomized and natural experiments," *Journal of Scandinavian Studies in Criminology and Crime Prevention* 18.2, pp. 210-222, 2017.

[2] A. Castillo, et al. "Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning." *Neurocomputing* 330, pp. 151-161, 2019.

[3] R. Olmos, S. Tabik, and F. Herrera. "Automatic handgun detection alarm in videos using deep learning." *Neurocomputing* 275, pp. 66-72, 2018.

[4] X. Wu, D. Sahoo, and S. CH Hoi. "Recent advances in deep learning for object detection." *Neurocomputing* 396, pp. 39-64, 2020.

[5] G.K. Verma, and A. Dhillon. "A handheld gun detection using faster r-cnn deep learning." *Proceedings of the 7th international conference on computer and communication technology*, pp. 84-88, 2017.

[6] M.P.J. Ashby. "The value of CCTV surveillance cameras as an investigative tool: An empirical analysis." *European Journal on Criminal Policy and Research* 23.3, pp. 441-459, 2017.

[7] S. Coşar, et al. "Toward abnormal trajectory and event detection in video surveillance." *IEEE Transactions on Circuits and Systems for Video Technology* 27.3, pp. 683-695, 2016.

[8] J. Park, et al. "A comparison of convolutional object detectors for real-time drone tracking using a PTZ camera." *2017 17th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2017.

[9] G. Mathur, and M. Bundele. "Research on intelligent video surveillance techniques for suspicious activity detection critical review." *2016 international conference on recent advances and innovations in engineering (ICRAIE)*. IEEE, pp.1-8, 2016.

[10] F. Gelana, and A. Yadav. "Firearm detection from surveillance cameras using image processing and machine learning techniques." *Smart innovations in communication and computational sciences*. Springer, Singapore, pp. 25-34, 2019.

[11] S. Ren, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28, 2015.

[12] J. Dai, et al. "R-fcn: Object detection via region-based fully convolutional networks." *Advances in neural information processing systems* 29, 2016.

[13] Z.Q. Zhao, et al. "Object detection with deep learning: A review." *IEEE transactions on neural networks and learning systems* 30.11, pp. 3212-3232, 2019.

[14] R.K. Tiwari, and G.K. Verma. "A computer vision based framework for visual gun detection using harris interest point detector." *Procedia Computer Science* 54, pp. 703-712, 2015.

[15] B. Cheng, et al. "Revisiting rcnn: On awakening the classification power of faster rcnn." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

[16] Y. Ren, C. Zhu, and S. Xiao. "Object detection based on fast/faster RCNN employing fully convolutional architectures." *Mathematical Problems in Engineering* 2018, 2018.

[17] J. Redmon, and A. Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767*, 2018.

[18] H. Kim, et al. "Detecting construction equipment using a region-based fully convolutional network and transfer learning." *Journal of computing in Civil Engineering* 32.2 : 04017082, 2018.